

REMARKS

I. Obviousness Rejection based on Huang and Coorman

Claims 1 to 2, 4 to 6, 8 to 9 and 11 to 12 were rejected under 35 U.S.C. 103 (a) over Huang, et al (referred to below as "Huang") and Coorman, et al, (referred to below as "Coorman").

Huang does describe a method of generating a description of multimedia data. The method is useful in an improved MPEG process of compression and coding of the multimedia data, namely MPEG-7. According to claim 12 of Huang a system is disclosed including a server for generating the description of multimedia data and storing it in a database and a client terminal coupled with the server for searching the data base and accessing the descriptions stored in the database.

Huang does describe the general features of the MPEG method for generating a description and its terms as well as a claimed improvement of MPEG (columns 5 to 7). In fact, applicants illustrate their inventive method with an example based on MPEG-7, as explained on pages 5 and following of their specification.

However, as admitted in the Office Action, Huang fails to teach or suggest including "a set of phonetic translation hints in the data stream of the multimedia data in addition to the textual description" (see claim 1).

Coorman is cited as providing that feature but it is respectfully submitted that Coorman also fails to teach or suggest including a set of phonetic translation

hints in the data stream of the multimedia data in addition to the textual description. The underlined portion of the previous sentence is the key.

It is respectfully submitted that a data stream of multimedia data in addition to the textual description means that the data stream is a coded stream of text data, such as computer-coded word processor generated digital data including text information comprising written words (i.e. "textural description" in claim 1). For example, page 5 of the specification defines "data" at least in the case of MPEG-7, as "audio-visual information that will be described using MPEG-7", i.e. a stream of computer-coded text data.

Coorman actually discloses a speech synthesizer for generating speech. The speech synthesizer of Coorman includes a large speech database of speech waveforms and a speech waveform selector that selects waveforms referenced by the speech database. In some embodiments a speech concatenator combines the various selected waveforms to form a speech signal (see column 4, line 50, to column 8, line 45).

However the input to the speech synthesizer is not a data stream of multimedia data (as claimed in applicants' claim 1 and 3). Instead it is "phonetic specifications" or "phonetic descriptors" or "polyphone designators" (column 8, line 61; column 9, line 4; column 4, line 53). These phonetic descriptors are produced according to the detailed description in a text processor 101 (column 8, line 65), which translates text data, such as word processor terms, into these phonetic descriptors. An example of this sort of translation is given in column 9, line 13 to 20, where Coorman shows how to translate the words "Hello,

Goodbye!" into phonetic descriptors.

The "data stream of multimedia data" is the stream of data that is fed into the text processor 101 in Coorman. The text data that is input to the text processor in Coorman is not described in detail. The subject matter in columns 9 and 10 of this reference solely describes the operation of the speech synthesizer including phonetic descriptors, the process or waveform selection by the waveform selector and the concatenation process. There is no discussion in Coorman of features of the data stream of multimedia data that are input to the text processor, which include phonetic descriptions or phonetic translation hints that would allow one to by-pass the transcription device that converts text to a phonetic description (which is one of the objects of the invention, according to page 3, line 13 to 16, of the specification).

Thus there is no hint or suggestion that the data stream of multimedia data that is presented or initially supplied for conversion to speech/audio signals contains "phonetic translation hints that specify the phonetic transcription of words of the textual description".

Applicants' specification explains how these phonetic translation hints are included in the initial text data stream in the case of MPEG-7. Multimedia data by definition includes a plurality of different data types, e.g. auditory data, visual data, printed data, so that XML must specify the data type prior to transmitting data of that particular type, so that the processor receiving the data has the ability to handle it appropriately. Applicants' improvement of the MPEG method involves setting up a new class or type of data, namely "Phonetic translation

hints". This requires that an identifier or header must signal the presence of the "phonetic translation hints" and then following this header or identifier the hints are provided and correlated with a word that they correspond to. This is clearly explained on page 11 of the applicants' specification.

There is no doubt regarding the meaning of the applicants' words in the amended claim 1 nor is their scope too broad. Claim 1 clearly and unambiguously states that the "phonetic translation hints" are embedded in the data stream of multimedia data in addition to the textual description to upgrade that data stream. When the "phonetic translation hint" comprises the phonetic description or transcription itself, the transcription can be directly input to the speech synthesizer and the text processor that generates the phonetic specification from the text data can be by-passed, thus shortening the process.

The applicants' method has several applications and associated benefits that result in clear improvements over the prior art. Phonetic transcription processors or text processors, such as processor 101, of Coorman are usually language specific, i.e. they generally assume that the text uses a single language, usually the native language of the user. Nevertheless multimedia presentations, for example a geography text, sometimes include words in a foreign language. Often the transcription processor does not produce the right pronunciation for the foreign language terms. However applicants' method solves this problem because a "phonetic translation hint" for the foreign word is included in the data stream of multimedia data that is supplied to the transcription or text processor, which may include the actual phonetic description (thus relieving the

text processor from the need to generate it).

There is not the slightest hint or suggestion of this method for updating a data stream of multimedia data in the Coorman reference, especially in columns 9 and 10. Coorman is only concerned with speech synthesis methods and speech synthesizers and is silent regarding fundamental changes in the data stream of multimedia data that is supplied, e.g. to the text processor.

It is well established by many U. S. Court decisions that to reject a claimed invention under 35 U.S.C. 103 there must be some hint or suggestion in the prior art of the modifications of the disclosure in a prior art reference or references used to reject the claimed invention, which are necessary to arrive at the claimed invention. For example, the Court of Appeals for the Federal Circuit has said:

"Rather, to establish obviousness based on a combination of elements disclosed in the prior art, there must be some motivation, suggestion or teaching of the desirability of making the specific combination that was made by the applicant... Even when obviousness is based on a single reference there must be a showing of a suggestion of motivation to modify the teachings of that reference.." *In re Kotzab*, 55 U.S.P.Q. 2nd 1313 (Fed. Cir. 2000). See also M.P.E.P. 2141

For the foregoing reasons and because of the changes in amended claim 1, withdrawal of the rejection of claims 1 to 2, 4 to 6, 8 to 9 and 11 to 12 under 35 U.S.C. 103 (a) over Huang, et al (referred to below as "Huang") and Coorman, et al, (referred to below as "Coorman") is respectfully requested.

II. Obviousness Rejection based on Huang, Coorman and Carter

Claim 3 was rejected as obvious under 35 U.S.C. 103 (a) over Huang, et al, (referred to below as "Huang") and Coorman, et al, (referred to below as "Coorman"), and further in view of Carter, et al, (referred to below as "Carter").

Claim 3 has now been amended so that it is now independent by including the features of claim 1 in it and wording has been changed somewhat to emphasize that the term "data stream" means the stream of multimedia data that may be inputted to a phonetic transcription device to convert encoded electronic signals including multimedia information, in part, into phonetic descriptions (naturally of only the audio or speech data present in the electronic signals).

Carter discloses a method and apparatus and computer program for reducing load on a text-to-speech converter in a message system capable of text to speech conversions of E-mail documents (title, abstract). Column 4, lines 1 to 23, does disclose a converter apparatus with a cache memory that stores certain text segments of a received E-mail message in a cache memory along with the converted speech signals for those text segments. Then when playback of other E-mail messages is requested if the other E-mail messages contained a text segment that is stored in the cache memory along with the speech signals the text-to-speech conversion process for that text segment is by-passed and the stored speech signal for it is used. For example, see figure 3 and the description associated with it.

However the applicants' method as claimed in claim 3 differs from and is not obvious from these disclosures in Carter, although both the method

described in Carter and that in claim 3 include features in their respective methods that permit faster more economical execution of the respective methods.

In the case of the disclosure of Carter actual speech signals are stored in the cache memory along with the associated text segment that may be repeated. In contrast in the case of the applicants' method, as claimed in claim 3, a phonetic transcription hint or a phonetic transcription, embedded in the data stream of multimedia data is simply not repeated when the associated text segment is repeated in the data stream.

Second, the method of applicants' claim 3 requires that the previous phonetic transcription hint is valid for a defined portion or all of the textual description in the data stream (claim 3). No storage and no retrieval of speech signals or any other data in and from a memory are required in the claimed method of applicants' claim 3 in contrast to the method of claim 5 of Carter, which is described in part in column 4.

Third, because storage and retrieval of speech signals in a memory is required in the method of Carter, the savings of conversion work when text segments are repeated is limited. The embodiments using the cache memory (for speed) is particularly limited, as explained in column 4, lines 13 to 23, of Carter because the cache can only function with repeated text segments that have 40 or fewer characters. There are no such limits to applicants' method. The phonetic translation hint could in principle be larger than 40 characters and apply to in principle to a large text segment, for example, an entire sentence or phrase in a

foreign language.

As to claims 4 and 5, in reference to both the rejections of claims 1 and 3 under 35 U.S.C. 103 (a), neither Coorman nor Carter provide any hint or suggestion of providing "phonetic translation hints" in an MPEG data stream of multimedia data for the purpose of providing a more accurate phonetic translation and/or by-passing the transcription process in which a textual description is converted to a phonetic description.

For the foregoing reasons and because of the changes in claim 3, withdrawal of the rejection of claim 3 as obvious under 35 U.S.C. 103 (a) over Huang, et al, and Coorman, et al, and further in view of Carter, et al, is respectfully requested.

III. Obviousness Rejection based on Huang, Coorman and Sharman

Claims 7 and 10 were rejected as obvious under 35 U.S.C. 103 (a) over Huang, et al, (referred to below as "Huang") and Coormann, et al, (referred to below as "Coormann"), and further in view of Sharman, et al (referred to below as "Sharman").

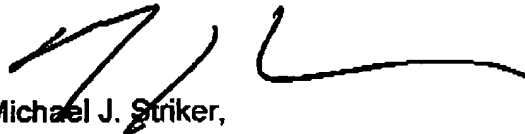
The features of claims 7 and 10 are currently not being relied on to establish patentability of the claimed method. Instead these features are features of preferred embodiments of the amended method claim 1.

For the foregoing withdrawal of the rejection of claims 7 and 10 as obvious under 35 U.S.C. 103 (a) over Huang, et al, and Coorman, et al, and further in view of Sharman, et al, is respectfully requested.

Should the Examiner require or consider it advisable that the specification, claims and/or drawing be further amended or corrected in formal respects to put this case in condition for final allowance, then it is requested that such amendments or corrections be carried out by Examiner's Amendment and the case passed to issue. Alternatively, should the Examiner feel that a personal discussion might be helpful in advancing the case to allowance, he or she is invited to telephone the undersigned at 1-631-549 4700.

In view of the foregoing, favorable allowance is respectfully solicited.

Respectfully submitted,



Michael J. Striker,

Attorney for the Applicants

Reg. No. 27,233